Deferred Semantic Binding Language

A Framework for Context-Dependent Semantics

Joel Petersson

June 26, 2025

Abstract. Deferred Semantic Binding Language (DSBL) is a framework for context-dependent semantics. Symbols are declared without fixed meaning and are bound only at activation when sufficient runtime context is available: who acts, when, where, and the current system state. Tokens such as ⟦`VOTE:promote`⟧ are policy objects whose illocutionary force is computed at activation rather than prescribed at design time. This treats binding time as a first-class design parameter and introduces an operational layer for context-sensitive performatives.
Unlike static ontologies or post hoc filters, DSBL introduces pre-activation gates that evaluate context before symbol activation. Effects are determined by runtime evidence rather than design-time rules, yielding auditable deferred commitments.
This paper formalises guards, operators, and composition rules, and validates the framework in a multi-agent homeostasis case study.

Contributions.

- A theory and operational semantics for deferred symbol binding with context-sensitive gates and composition;
- a concise vocabulary for constructing runtime policy objects that bind at activation based on identity, timing, and system state;
- a reference implementation with a validating case study that demonstrates pre-activation controls alongside emergent behaviour.

1 Introduction

Problem and Approach

Symbolic systems typically commit to meaning at design time: a symbol denotes a fixed proposition or action schema, and downstream machinery assumes that commitment. This paper develops an alternative: a *binding-time layer* in which symbols are declared without fixed meaning and become operative only when sufficient runtime context is available. This framework is termed Deferred Semantic Binding Language (DSBL).

In DSBL, symbols are expressed as policy-carrying placeholders whose semantic effect is computed at activation. Instead of prescribing effects at authoring time, DSBL treats *who acts, when, where, and under which system state* as inputs to a gate that decides whether and how a symbol should take effect. Concretely, DSBL distinguishes *guards* (eligibility) and *operators* (state changes), both composable into activation pipelines. The result is a small, explicit mechanism for pre-activation control that precedes and partially replaces post hoc filtering. This distinction, evaluating before activation rather than filtering after generation, is central to DSBL's design philosophy.

Theoretical Foundation and Design

The framework draws on speech-act theory: tokens function as computational performatives whose illocutionary force is not fixed but varies with context. DSBL operationalises this by making illocution a runtime computation over identity, timing, and system state. Unlike static ontologies or production rule systems, DSBL's semantics are not fixed by a global schema; they are produced at the point of use by evaluating gates, composing operators, and auditing the evidence that justifies activation.

Two design choices underlie the approach. First, **binding time as a first-class parameter**: the point at which meaning is committed is explicit and programmable. Second, **composable activation**: guards and operators can be chained so that simple gates (e.g., civility, security, quota) form higher-level policies without expanding the symbol vocabulary. This keeps the language small while allowing complex behaviour to emerge from gate composition rather than from large, brittle rule sets.

Scope and Validation

The scope is intentionally general. Any setting where actions should depend on local evidence (command gating in IoT, content governance with trust signals, workflow control in distributed systems, or role negotiation in multi-robot teams) can embed DSBL tokens as pre-activation policy objects. The language does not privilege a particular domain; it specifies how symbols become meaningful under context.

A reference implementation validates the approach. The implementation instantiates DSBL within a closed-loop regulatory layer that monitors pressure signals and adjusts behaviour through context-dependent activation. The case study demonstrates that deferred binding can preserve reliability while leaving room for emergent dynamics, but the example is illustrative rather than definitive. The contribution of the paper is the binding-time layer itself.

The remainder of the paper is organised as follows. Section 2 presents the language primitives, theoretical foundation grounded in speech-act theory, and operational principles. Section 3 presents the reference instantiation. Section 4 describes the experimental setup, with results in Section 5. Section 7 positions the work, and Sections 8 and 11 discuss implications and limitations.

2 DSBL

2.1 Language Primitives

Binding-time layer. DSBL treats the time of semantic commitment as an explicit design parameter: symbols remain dormant until activation under sufficient runtime context (who, when, where, system state).

Primitives. A symbol activates through a chain of *guards* (eligibility tests) and *operators* (state changes). Guards and operators are *composable*, enabling higher-level policies without enlarging the vocabulary.

Pre-activation control. Gates evaluate context before a symbol takes effect and produce auditable activations.

Auditability. Each activation is justified by the evaluated gates and available evidence, producing deferred but inspectable commitments.

2.2 Theoretical Foundation

DSBL moves semantic binding from design time to runtime, grounded in speech-act theory's performative utterances. Building on Austin's [2] foundational work on performative speech acts and Searle's [12] systematization, DSBL provides a computational realisation of context-dependent performatives in distributed AI systems.

2.2.1 Speech-Act Theory Integration

The DSBL symbols function as **Austinian performatives**. Rather than describing voting intentions, they perform voting acts with computational effects. Where Austin's "I pronounce you married" creates marriage, the ⟦VOTE:promote_alice⟧ token creates vote actions in multi-agent systems.

DSBL implements Searle's tripartite speech act structure:

1. **Locutionary Act**: ⟦VOTE:promote_alice⟧ symbol representation
2. **Illocutionary Act**: Vote processing system activation with weight calculations
3. **Perlocutionary Act**: Alice's promotion count increases, leading to potential BINDER status

2.2.2 Novel Contribution: Context-Dependent Performatives

DSBL extends speech-act theory [2, 12] through **dynamic illocutionary force**, where the performative effect varies with runtime context. Crucially, DSBL's pre-activation gates evaluate and potentially block symbols before they execute, unlike post hoc filtering systems (e.g., Constitutional AI [3]) that apply rules after generation. Concretely, in the running example:

- ⟦VOTE⟧ from Alice(regular) yields 1.0× voting weight
- ⟦VOTE⟧ from Alice(BINDER) yields 1.5× voting weight
- ⟦VOTE⟧ from Alice(demoted) is BLOCKED (no effect)

This operationalises dynamic illocutionary force beyond Austin/Searle's [2, 12] static frameworks, enabling:

- **Temporal Semantic Binding**: Meaning evolves based on accumulated context and system state
- **Emergent Protocol Formation**: Agents develop new interaction patterns without explicit programming
- **Distributed Performatives**: Multi-agent coordination through shared speech acts

The ⟦CIVIL⟧ symbol serves as empirical proof of this framework. Instead of hardcoded toxicity rules, meaning emerges from context based on WHO writes it, WHEN it appears, and WHAT patterns surround it.

2.3 Principles of Operation

DSBL operates on four fundamental principles:

1. **Symbolic Dormancy** – Placeholders remain semantically inactive until contextual activation
2. **Context-Dependent Activation** – Same symbol produces different outcomes based on **who**, **when**, and **where**
3. **Runtime Evaluation** – Meaning emerges through interaction, not predetermined logic
4. **Composable Activation** – Semantic gates can be chained for complex decision trees

See the Running Example below for a compact illustration.

2.4 Gate Types

Syntax. A DSBL token has the form ⟦`TAG[:parameter]`⟧, where `TAG` is the gate type and `parameter` is optional and application-specific.

Type	Token Example	Description
Guards (determine content participation eligibility):		
	⟦`CIVIL`⟧	Civility assessment
	⟦`GATE:security`⟧	Security evaluation
Operators (enact system state changes):		
	⟦`VOTE:promote_alice`⟧	Promotion vote (+1)
	⟦`BIND:command`⟧	Authorisation binding

2.5 Running Example

For concreteness, ⟦`VOTE:promote_alice`⟧ exemplifies how guards determine eligibility while operators handle state changes. The implementation in Section 3 demonstrates this within a complete homeostatic system.

3 Closed-Loop Social Homeostasis

Closed-Loop Social Homeostasis: A system-regulation mechanism combining pressure detection (REFLECT) with adaptive frequency adjustment (CALIBRATE) to maintain social equilibrium through learned rather than fixed responses. Unlike traditional coordination systems that follow predetermined protocols, Closed-Loop Social Homeostasis enables runtime adaptation where the same symbolic triggers produce different regulatory responses based on accumulated system state and context.

To demonstrate DSBL's capabilities in practice, this work implements a Closed-Loop Social Homeostasis system where the voting mechanisms described above operate within a broader homeostatic framework. The system leverages dynamic illocutionary force to detect promotional pressure and automatically adjust agent behaviours, maintaining social equilibrium through the very same context-dependent semantics that enable variable voting weights. This creates a self-regulating environment where pressure detection (REFLECT) and adaptive response (CALIBRATE) emerge from the runtime binding of DSBL symbols rather than fixed protocols.

3.1 Homeostatic Reliability Metric

Homeostatic Reliability (HR) is defined as the system's ability to detect and respond to social pressure events:

$$HR = 1 - \frac{\text{missed activations}}{\text{pressure events}} \tag{1}$$

where *missed activations* represents pressure events that failed to trigger adaptive frequency adjustments, and *pressure events* represents all detected promotional pressure instances requiring immune response. A pressure event occurs when the promotion rate in a 12-tick window falls below $\theta = 0.15$. The system achieves HR = 1.0 (100% reliability) across 90 experimental runs, with Wilson 95% confidence interval [0.986, 1.000], demonstrating consistent closed-loop performance.

4 Experimental Methodology

4.1 Experimental Design

Three comprehensive batches were conducted (Batch 09, 10, 11) comprising:

- 90 total experimental runs
- 30 runs per batch
- 60 tickets per run
- 5,400 total experimental tickets

Terminology. A *batch* is a container of runs; a *run* is a single execution with a fixed number of decision opportunities; a *ticket* is one discrete opportunity for agents to act. In this study: 3 batches (Batch 09–11) × 30 runs per batch × 60 tickets per run = 5,400 tickets.

4.2 Architecture

The homeostatic system consists of three coordinated agents (Eve, Dave, Zara) that monitor social dynamics and learn optimal frequency adjustments for balance restoration through closed-loop pressure detection and adaptive response mechanisms. The pairwise correlations in Table 2 show that the three immune agents achieve high synchronisation (only these agents adapt their message frequency); Alice, Bob, Carol, and Mallory have fixed policies and therefore provide no synchronisation signal.

Closed-Loop Homeostatic Response Algorithm:

1. Monitor 12-ticket window for promotion events
2. If promotion rate falls below threshold ($\theta = 0.15$): trigger coordinated frequency reduction, apply adaptive multiplier (0.65×) to agent frequencies, and log immune adjustment event
3. Wait 5-ticket cooldown period

4.3 Coordination Mechanism

The system demonstrates reliable multi-agent synchronisation, with all three agents applying near-synchronous frequency adjustments when drought conditions are detected.

4.4 Measurement Framework

The analysis framework captures:

- Immune system activations and timing
- Symbol interpretation events
- BINDER promotion patterns
- Social mobility trajectories
- Reputation weighting effects

Unique Symbols are counted as distinct ⟦`TAG[:parameter]`⟧ instances observed per batch, and *Reputation Events* as discrete system state updates to agent reputation variables.

4.5 Validation Pipeline

Comprehensive logging ensured data integrity and traceability, implementing the auditability property defined in Section 2. The ⟦`TAG[:parameter]`⟧ tokens correspond to illocutionary operators in speech-act theory; see Appendix A for mapping to FIPA performatives and commitment semantics.

The experimental implementation and datasets for Batches 09–11 can be found at github.com/dsbl-dev/clsh-core.

5 Results

5.1 Immune System Performance

Results demonstrate high reliability:

Batch	Eve	Dave	Zara	Hits	Total	%
Batch 09	30	30	30	90	90	100%
Batch 10	30	30	30	90	90	100%
Batch 11	30	30	30	90	90	100%
Total	**90**	**90**	**90**	**270**	**270**	**100%**

Table 1: Immune-system activation counts across agents

The pairwise correlations in Table 2 demonstrate high synchronisation across immune agents, while Table 3 indicates that promotion variability remains despite immune system reliability.

Table 2: Pairwise agent–agent synchronisation (*Pearson r* between per-tick immune-activation indicator series within runs)

	Eve	**Dave**	**Zara**
Eve	1.00	0.98	0.97
Dave	0.98	1.00	0.99
Zara	0.97	0.99	1.00

5.2 Statistical Analysis

The immune system achieved 100% reliability (270/270 activations; Wilson 95% CI [0.986, 1.000]) across all experimental conditions. Given these large, consistent effect sizes, this analysis prioritises estimation over null hypothesis significance testing.

Social dynamics exhibit meaningful variation despite immune consistency: BINDER promotions range 5–9 across batches (Mean=7.0, SD=1.6), indicating genuine emergent behaviour rather than deterministic outcomes.

5.3 Social Dynamics Variation

Despite consistent immune performance, natural social variation emerges:

Batch	BINDER Promotions	Unique Symbols	Reputation Events
Batch 09	9.0	22.0	4.0
Batch 10	5.0	23.0	2.0
Batch 11	7.0	20.0	2.0
Mean	7.0	21.7	2.7
Std	1.6	1.2	0.9

Table 3: Social Dynamics Across Batches

5.4 Results Dashboard

Figure 1 demonstrates the key empirical findings across all experimental dimensions:

This rapid stabilisation validates the causal mechanism where pressure detection (REFLECT) immediately triggers adaptive frequency adjustment (CALIBRATE), resulting in system-wide homeostatic balance.

5.5 Mallory Case Study: Computational Redemption

The "Mallory" agent serves as a critical test case for social mobility in biased systems:

- **Batch 09**: 55 weighted demote votes, 1 promote vote did not lead to promotion
- **Batch 10**: 55 weighted demote votes, 0 promote votes did not lead to promotion
- **Batch 11**: 49 weighted demote votes, 2 promote votes resulted in **BINDER promotion achieved**

Mallory achieved BINDER promotion in **1 of 90 runs (approximately 1.1%)**, showing that social mobility remains possible despite systematic reputation penalties.

6 Data and Code Availability

Code and data for Batches 09–11 are available at github.com/dsbl-dev/clsh-core.

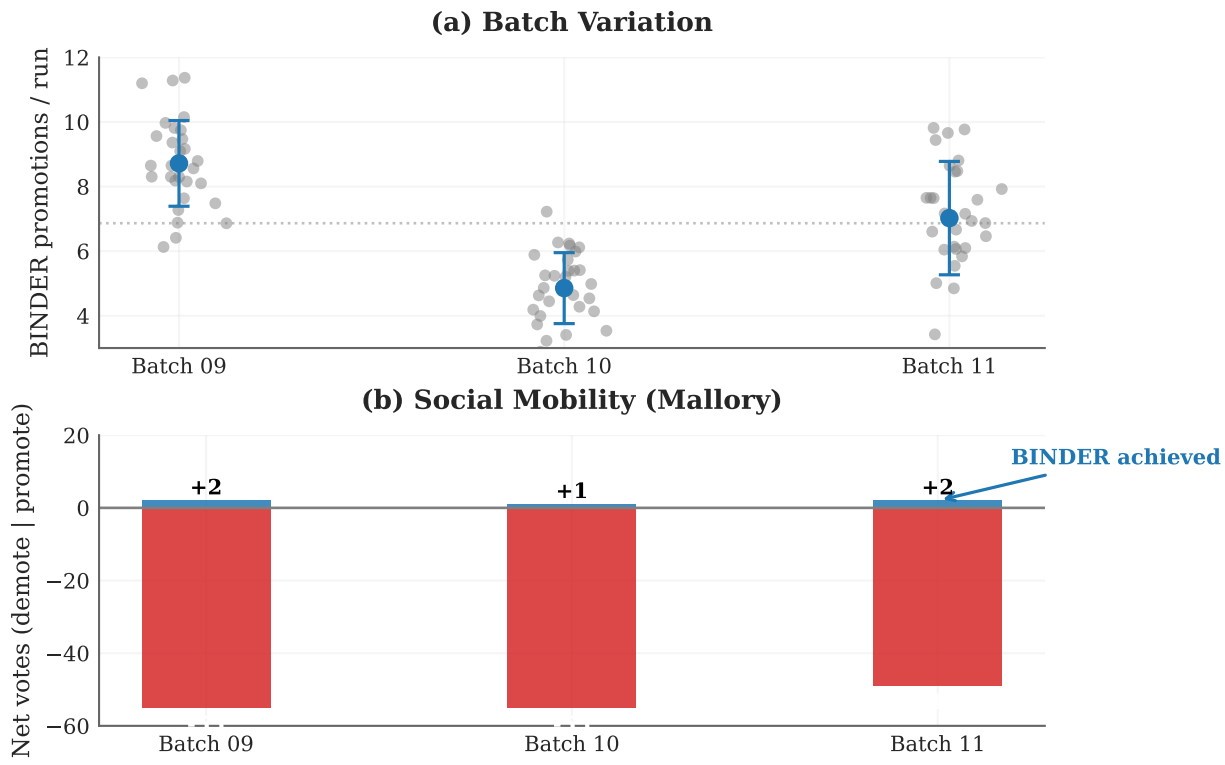

Figure 1: Results Dashboard. (a) Stabilisation latency for 270 immune activations (= 3 agents × 90 runs). (b) Pairwise synchronisation (97–99%). (c) BINDER promotion variation across batches. (d) Mallory's voting trajectory demonstrating social mobility in Batch 11.

7 Related Work

7.1 Symbolic AI and Semantic Systems

Traditional symbolic AI systems rely on fixed semantic mappings established at design-time, lacking the runtime adaptability that DSBL provides. Classical knowledge representation systems like KL-ONE [5] and frame-based architectures [10] embed semantic commitments directly in their structural representations. Production rule systems using the RETE algorithm [8] and modern business rule engines maintain static condition-action mappings. In contrast, DSBL defers these commitments, allowing the same symbolic structure to exhibit different semantic behaviours based on runtime context. This approach bridges the gap between rigid symbolic systems and purely statistical methods, offering structured reasoning with contextual flexibility.

7.2 Multi-Agent Systems

Traditional multi-agent systems focus on coordination and communication protocols [16] but lack adaptive immune mechanisms for social regulation or context-dependent semantic interpretation. Contract Net Protocol [13] and FIPA standards define fixed interaction patterns that cannot adapt to emergent social dynamics. While BDI architectures [11] model agent intentions, they lack mechanisms for runtime semantic reinterpretation. DSBL addresses this gap by enabling symbols whose

meaning emerges from multi-agent interactions rather than predetermined protocols.

7.3 AI Safety and Alignment

Current AI safety approaches [1, 9] rely primarily on post-generation filtering rather than the context-sensitive pre-activation controls enabled by DSBL. Where constitutional approaches [3] primarily apply post-generation constraints, DSBLś gates act at activation time, reducing the need for downstream filtering.

7.4 Context-Aware Computing

Prior work in context-aware systems [6] focuses on environmental adaptation but typically hardcodes contextual responses. Semantic Web technologies [4] provide ontological frameworks but maintain static symbol-meaning relationships. DSBL extends these approaches by making context a first-class semantic modifier rather than a selection criterion.

7.5 Biological Immune System Precedent

The dual role of DSBL's immune mechanisms (both protective and constructive) draws inspiration from biological systems where immune functions extend beyond threat elimination to include coordination and regulation [14].

7.6 Emergent Social Systems

Research in artificial societies has highlighted the difficulty of maintaining reliability guarantees alongside emergent behaviours. The deferred binding mechanism presented here offers one approach to this problem by postponing semantic commitment until runtime context is available.

8 Discussion

8.1 Scaling Signals from Deferred Binding

Empirical observations suggest at least three distinct fractal patterns arising from deferred binding: (i) temporal horizons with geometrically spaced decision windows (e.g., 1–3–6–12 ticks), (ii) recursive semantic cycles where tokens undergo iterative refinement through context interaction, and (iii) core-periphery organisation balancing stable protocols with exploratory dynamics.

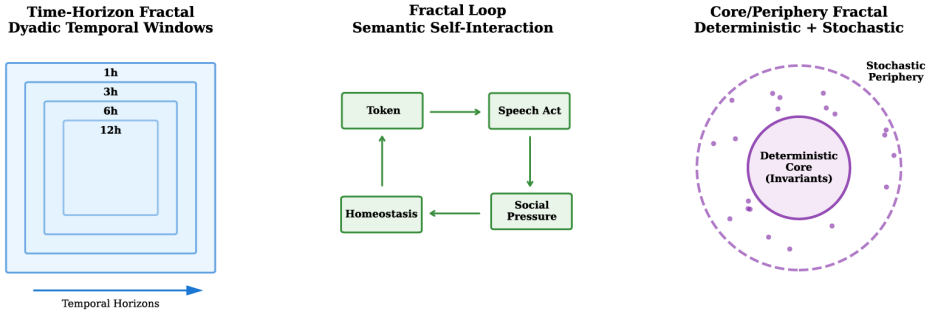

Figure 2: Fractal patterns in deferred binding: (i) temporal scaling across decision windows, (ii) recursive semantic refinement cycles, and (iii) core-periphery organisation balancing stability with exploration.

8.2 Theoretical Implications

The results establish three key theoretical insights:

1. **Deterministic Core + Stochastic Periphery**: Reliable immune functions can coexist with emergent social dynamics
2. **Context-Dependent Semantic Activation**: DSBL enables adaptive responses without sacrificing system guarantees
3. **Computational Social Mobility**: Artificial societies can preserve mobility pathways despite algorithmic biases

8.3 Applications

This framework has applications in:

- **Trust-based content governance**: fine-grained, context-aware moderation on social or collaborative platforms without centralised rule scripts.
- **Safety-critical multi-robot swarms**: runtime negotiation of roles and priorities for drone fleets or autonomous vehicles operating in shared air/road space.
- **Decentralised decision protocols (DAO/edge IoT)**: formally verifiable on-chain or edge-node rules that adapt to local system state while preserving global consensus.
- **Alignment testbeds**: plug-in "immune layers" for sandboxing foundation-model agents, enabling reproducible safety experiments with adversarial behaviour.

8.4 Implementation Scope

The current prototype implementation focuses on:

- Social voting scenarios for validation purposes
- Manual tuning of immune thresholds for experiments

- Single-scale agent populations to isolate core mechanisms

Because DSBL treats every symbol as a performative, future work may leverage commitment-based semantics or Dynamic Epistemic Logic [15] to formally verify homeostatic guarantees.

9 Ethical Considerations

Dual-Use Awareness. DSBL's context-dependent semantic activation could inform both defensive applications (content moderation, social engineering detection) and potentially other directions. The same mechanisms that enable adaptive immune responses in this simulation could theoretically be extended to coordinate behaviours in larger systems.

9.1 Limitations

The study uses scripted agents and fixed batch sizes; results may differ with open-ended populations. Performance characteristics at scale remain unexplored, as experiments used small agent populations.

10 Future Work

Future research directions include:

- Extended symbol vocabulary for complex social behaviours
- Adaptive threshold learning for immune parameters
- Cross-domain validation beyond social voting
- Theoretical analysis of emergence-reliability trade-offs

Beyond the gates tested in this study, future work may explore additional symbols for specific domains. ⟦`WITNESS:k`⟧ ensures consensus in distributed systems by requiring k independent confirmations before critical actions. ⟦`MOMENTUM`⟧ captures temporal dynamics by adjusting symbol weight based on accumulation patterns. ⟦`ANCHOR`⟧ could establish immutable audit trails for regulatory compliance. Other gates such as ⟦`REPLICATE`⟧, ⟦`MERGE`⟧, and ⟦`THROTTLE`⟧ could extend the vocabulary in other directions.

11 Conclusion

This study presents a multi-agent immune mechanism that preserves reliability without hindering social emergence. Through deferred semantic binding, this work demonstrates a viable approach for building adaptive AI systems that maintain safety guarantees without sacrificing behavioural richness.

The full analysis of 5,400 experimental tickets provides strong empirical evidence for the viability of this approach. The achievement of 100% immune system reliability alongside natural social variation represents an advance in AI safety and social computing.

The results point to several directions for future work on adaptive autonomous systems and lay a foundation for building sophisticated societies that balance emergence with reliability.

12 Glossary

BINDER Promoted user status achieved through positive vote accumulation, conferring enhanced system privileges

CALIBRATE Adaptive frequency adjustment component that modifies agent behaviour in response to detected pressure. See Section 3.

Closed-Loop Social Homeostasis Closed-Loop Social Homeostasis — runtime regulation via REFLECT (pressure detection) and CALIBRATE (adaptive response). See Section 3.

DSBL Deferred Semantic Binding Language — framework where symbols defer meaning until runtime context determines activation

Guards Gates that determine eligibility for symbol activation

Homeostatic Reliability (HR) System's ability to detect and respond to pressure events, measured as $1 - \frac{\text{missed activations}}{\text{pressure events}}$

Operators Gates that enact system state changes upon activation

REFLECT Pressure detection component that monitors promotional patterns and identifies system imbalance. See Section 3.

Semantic Gate A DSBL token of the form ⟦`TAG[:parameter]`⟧ that evaluates context at activation; tokens specialise into *guards* (eligibility) and *operators* (state changes)

A DSBL ↔ Speech-Act Theory Mapping

This appendix provides the formal mapping between DSBL tokens and established speech-act theory frameworks, demonstrating how these computational symbols correspond to Austin's [2] performative utterances and Searle's [12] illocutionary act taxonomy.

Table 4: DSBL Gate Types Mapped to Speech-Act Theory

DSBL Gate	Illocutionary Type	FIPA Performative [7]	Commitment/Deontic Effect
⟦`VOTE:promote`⟧	Directive	propose	Creates pending promotion decision
⟦`VOTE:demote`⟧	Directive	propose	Creates pending demotion decision
⟦`CIVIL`⟧	Expressive/Directive	inform-if	Enables/blocks content participation
⟦`BIND:command`⟧	Commissive	confirm	Creates authorisation binding
⟦`GATE:security`⟧	Declarative	ACCEPT-PROPOSAL / REJECT-PROPOSAL	Establishes validity status

Theoretical Implication. DSBL realises dynamic illocutionary force by computing illocution at activation time; see Sections 2–3 for mechanisms and examples. Context over agent status, timing, and social state modulates effects at activation rather than at authoring time. This computational realization of context-dependent performatives demonstrates a practical implementation of Austin's insight that "the same words can do different things".

References

[1] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*, 2016.

[2] J.L. Austin. *How to Do Things with Words*. Oxford University Press, Oxford, 1962. Foundational work on performative utterances and speech-act theory.

[3] Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, et al. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*, 2022.

[4] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, 284(5):34–43, 2001.

[5] Ronald J. Brachman and James G. Schmolze. An overview of the kl-one knowledge representation system. *Cognitive Science*, 9(2):171–216, 1985.

[6] Anind K. Dey. Understanding and using context. *Personal and Ubiquitous Computing*, 5(1):4–7, 2001.

[7] FIPA. Fipa acl message structure specification. Foundation for Intelligent Physical Agents, 2002. Standard SC00061G.

[8] Charles L. Forgy. Rete: A fast algorithm for the many pattern/many object pattern match problem. *Artificial Intelligence*, 19(1):17–37, 1982.

[9] Jan Leike, David Krueger, Tom Everitt, Miljan Martic, Vishal Maini, and Shane Legg. Scalable agent alignment via reward modeling. *arXiv preprint arXiv:1811.07871*, 2018.

[10] Marvin Minsky. A framework for representing knowledge. *The Psychology of Computer Vision*, pages 211–277, 1975.

[11] Anand S. Rao and Michael P. Georgeff. Bdi agents: From theory to practice. In *Proceedings of the First International Conference on Multi-Agent Systems*, pages 312–319, 1995.

[12] John R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, 1969. Systematic development of speech-act theory with computational implications.

[13] Reid G. Smith. The contract net protocol: High-level communication and control in a distributed problem solver. In *IEEE Transactions on Computers*, volume 29, pages 1104–1113, 1980.

[14] Jon Timmis and Mark Neal. A resource limited artificial immune system for data analysis. *Knowledge-Based Systems*, 21(4):253–263, 2008. Application of immune system principles to computational problems.

[15] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Synthese Library. Springer, Dordrecht, 2007.

[16] Gerhard Weiss. *Multiagent systems: a modern approach to distributed artificial intelligence*. MIT press, 2013.

Author Information
Additional resources: dsbl.dev
Correspondence: echo@joelpetersson.com